

# Images\_and\_LinkedIn\_Profile(Data\_Analysis)

Allama Ikbal Sijan

2023-09-12

## Data Wrangling & Cleaning

```
df <- read_csv("Data_Clean.csv")
df <- df %>%
  select(!c(StartDate, EndDate, IPAddress, Status,
            RecordedDate, LocationLatitude, LocationLongitude, DistributionChannel,Q1))
```

```
df <- df %>% slice(-2)
```

```
asian_man_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

asian_man_df <- asian_man_df %>%
  filter(Asian_Man_1 != "N/A")

colnames(asian_man_df) <- asian_man_df[1,]
asian_man_df <- asian_man_df[-1, ]
colnames(asian_man_df)[7] = "Attention_check"
```

```
asian_woman_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

asian_woman_df <- asian_woman_df %>%
  filter(Asian_Woman_1 != "N/A")

colnames(asian_woman_df) <- asian_woman_df[1,]
asian_woman_df <- asian_woman_df[-1, ]
colnames(asian_woman_df)[7] = "Attention_check"
```

```
black_man_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

black_man_df <- black_man_df %>%
  filter(Black_Man_1 != "N/A")

colnames(black_man_df) <- black_man_df[1,]
black_man_df <- black_man_df[-1, ]
colnames(black_man_df)[7] = "Attention_check"
```

```
black_woman_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

black_woman_df <- black_woman_df %>%
  filter(Black_Woman_1 != "N/A")

colnames(black_woman_df) <- black_woman_df[1,]
black_woman_df <- black_woman_df[-1, ]
colnames(black_woman_df)[7] = "Attention_check"
```

```
blight_man_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

blight_man_df <- blight_man_df %>%
  filter(Blight_Man_1 != "N/A")

colnames(blight_man_df) <- blight_man_df[1,]
blight_man_df <- blight_man_df[-1, ]
colnames(blight_man_df)[7] = "Attention_check"
```

```
blight_woman_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

blight_woman_df <- blight_woman_df %>%
  filter(Blight_Woman_1 != "N/A")

colnames(blight_woman_df) <- blight_woman_df[1,]
blight_woman_df <- blight_woman_df[-1, ]
colnames(blight_woman_df)[7] = "Attention_check"
```

```
wasian_man_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

wasian_man_df <- wasian_man_df %>%
  filter(Wasian_Man_1 != "N/A")

colnames(wasian_man_df) <- wasian_man_df[1,]
wasian_man_df <- wasian_man_df[-1, ]
colnames(wasian_man_df)[7] = "Attention_check"
```

```
wasian_woman_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

wasian_woman_df <- wasian_woman_df %>%
  filter(Wasian_Woman_1 != "N/A")

colnames(wasian_woman_df) <- wasian_woman_df[1,]
wasian_woman_df <- wasian_woman_df[-1, ]
colnames(wasian_woman_df)[7] = "Attention_check"
```

```

white_man_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

white_man_df <- white_man_df %>%
  filter(White_Man_1 != "N/A")

colnames(white_man_df) <- white_man_df[1,]
white_man_df <- white_man_df[-1, ]
colnames(white_man_df)[7] = "Attention_check"

white_woman_df <- df %>%
  select(ResponseId, Column1, `Duration (in seconds)`, Progress, UserLanguage, Finished, Attention_check)

white_woman_df <- white_woman_df %>%
  filter(White_Woman_1 != "N/A")

colnames(white_woman_df) <- white_woman_df[1,]
white_woman_df <- white_woman_df[-1, ]
colnames(white_woman_df)[7] = "Attention_check"

new_df <- rbind(asian_man_df, asian_woman_df,
               black_man_df, black_woman_df,
               blight_man_df, blight_woman_df,
               wasian_man_df, wasian_woman_df,
               white_man_df, white_woman_df)

new_df = new_df[,!(names(new_df) %in% c("Timing - First Click","Timing - Last Click"))]

colnames(new_df)[8] <- "timing_page_submit_LP"
colnames(new_df)[9] <- "click_count_LP"

colnames(new_df)[10] <- "item1_leader"
colnames(new_df)[11] <- "timing_leader"
colnames(new_df)[12] <- "click_count_leader"

colnames(new_df)[13] <- "item2_independent"
colnames(new_df)[14] <- "timing_independent"
colnames(new_df)[15] <- "click_count_independent"

colnames(new_df)[16] <- "item3_ambitious"
colnames(new_df)[17] <- "timing_ambitious"
colnames(new_df)[18] <- "click_count_ambitious"

colnames(new_df)[19] <- "item4_loyal"
colnames(new_df)[20] <- "timing_loyal"
colnames(new_df)[21] <- "click_count_loyal"

colnames(new_df)[22] <- "item5_sensitive"
colnames(new_df)[23] <- "timing_sensitive"
colnames(new_df)[24] <- "click_count_sensitive"

colnames(new_df)[25] <- "item6_warm"
colnames(new_df)[26] <- "timing_warm"

```

```

colnames(new_df)[27] <- "click_count_warm"

colnames(new_df)[28] <- "item7_compassionate"
colnames(new_df)[29] <- "timing_compassionate"
colnames(new_df)[30] <- "click_count_compassionate"

colnames(new_df)[31] <- "item8_adaptable"
colnames(new_df)[32] <- "timing_adaptable"
colnames(new_df)[33] <- "click_count_adaptable"

colnames(new_df)[34] <- "item9_sincere"
colnames(new_df)[35] <- "timing_sincere"
colnames(new_df)[36] <- "click_count_sincere"

colnames(new_df)[37] <- "item10_reliable"
colnames(new_df)[38] <- "timing_reliable"
colnames(new_df)[39] <- "click_count_reliable"

colnames(new_df)[40] <- "item11_truthful"
colnames(new_df)[41] <- "timing_truthful"
colnames(new_df)[42] <- "click_count_truthful"

colnames(new_df)[43] <- "item12_race"
colnames(new_df)[44] <- "timing_race"
colnames(new_df)[45] <- "click_count_race"

colnames(new_df)[46] <- "dem1_age"
colnames(new_df)[47] <- "dem2_gender"
colnames(new_df)[48] <- "dem3_ethnicity"
colnames(new_df)[49] <- "dem3_ethnicity_other"

colnames(new_df)[50] <- "dem4_employment_status"
colnames(new_df)[51] <- "dem5_occupation"

```

```

new_df <- new_df %>%
  filter(Finished == TRUE & Attention_check == "Digital Marketing Role")

```

```

fac_var <- c("Group", "dem2_gender", "dem3_ethnicity", "dem4_employment_status")
new_df <- new_df %>%
  mutate(across(fac_var, as.factor))

```

```

num_var <- c("Duration (in seconds)", "timing_page_submit_LP", "click_count_LP", "item1_leader", "timing_page_submit_LP")
new_df <- new_df %>%
  mutate(across(num_var, as.numeric))

```

## Descriptive statistics

```
describe(new_df)
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to min; returning Inf
```

```
## Warning in FUN(newX[, i], ...): no non-missing arguments to max; returning -Inf
```

##	vars	n	mean	sd	median	trimmed	mad	min
## Response ID*	1	608	304.50	175.66	304.50	304.50	225.36	1.00
## Group*	2	608	5.48	2.86	5.00	5.48	2.97	1.00
## Duration (in seconds)	3	608	209.88	179.03	162.00	175.32	69.68	43.00
## Progress*	4	608	1.00	0.00	1.00	1.00	0.00	1.00
## User Language*	5	608	1.00	0.00	1.00	1.00	0.00	1.00
## Finished*	6	608	1.00	0.00	1.00	1.00	0.00	1.00
## Attention_check*	7	608	1.00	0.00	1.00	1.00	0.00	1.00
## timing_page_submit_LP	8	608	42.06	50.92	28.66	31.96	19.24	1.41
## click_count_LP	9	608	1.78	4.02	0.00	0.85	0.00	0.00
## item1_leader	10	608	5.74	1.05	6.00	5.85	1.48	1.00
## timing_leader	11	608	11.76	14.70	8.35	9.29	4.92	1.64
## click_count_leader	12	608	2.18	1.92	1.00	1.77	0.00	0.00
## item2_independent	13	608	5.81	1.12	6.00	5.95	1.48	1.00
## timing_independent	14	608	8.12	12.85	5.44	6.02	2.63	0.94
## click_count_independent	15	608	1.88	1.45	1.00	1.58	0.00	0.00
## item3_ambitious	16	608	6.14	1.03	6.00	6.30	1.48	1.00
## timing_ambitious	17	608	8.08	19.04	4.59	5.04	2.02	0.99
## click_count_ambitious	18	608	1.79	1.25	1.00	1.54	0.00	0.00
## item4_loyal	19	608	5.13	1.21	5.00	5.14	1.48	1.00
## timing_loyal	20	608	7.55	9.98	4.97	5.58	2.33	1.30
## click_count_loyal	21	608	1.85	1.41	1.00	1.56	0.00	0.00
## item5_sensitive	22	608	5.13	1.15	5.00	5.12	1.48	1.00
## timing_sensitive	23	608	10.42	50.75	4.82	5.38	2.15	1.23
## click_count_sensitive	24	608	1.88	1.41	1.00	1.61	0.00	0.00
## item6_warm	25	608	5.10	1.20	5.00	5.14	1.48	1.00
## timing_warm	26	608	6.59	12.51	4.15	4.47	1.75	1.18
## click_count_warm	27	608	1.83	1.41	1.00	1.55	0.00	0.00
## item7_compassionate	28	608	5.04	1.16	5.00	5.04	1.48	1.00
## timing_compassionate	29	608	5.72	9.75	3.77	4.08	1.51	1.03
## click_count_compassionate	30	608	1.77	1.40	1.00	1.51	0.00	0.00
## item8_adaptable	31	608	5.83	1.00	6.00	5.93	1.48	1.00
## timing_adaptable	32	608	5.99	9.61	3.90	4.20	1.56	1.03
## click_count_adaptable	33	608	1.75	1.26	1.00	1.52	0.00	0.00
## item9_sincere	34	608	5.44	1.17	6.00	5.51	1.48	1.00
## timing_sincere	35	608	6.51	15.15	3.69	3.98	1.48	0.85
## click_count_sincere	36	608	1.74	1.22	1.00	1.49	0.00	0.00
## item10_reliable	37	608	6.03	0.99	6.00	6.16	1.48	2.00
## timing_reliable	38	608	6.11	12.85	3.60	3.95	1.55	0.82
## click_count_reliable	39	608	1.75	1.20	1.00	1.52	0.00	0.00
## item11_truthful	40	608	5.65	1.11	6.00	5.74	1.48	1.00
## timing_truthful	41	608	7.07	32.02	3.51	3.85	1.41	0.94
## click_count_truthful	42	608	1.75	1.17	1.00	1.53	0.00	0.00
## item12_race	43	0	NaN	NA	NA	NaN	NA	Inf
## timing_race	44	608	15.65	40.67	10.36	11.48	6.13	1.86
## click_count_race	45	608	2.23	2.35	1.00	1.73	0.00	0.00
## dem1_age	46	608	37.25	8.14	36.00	37.25	8.90	20.00
## dem2_gender*	47	608	2.58	0.52	3.00	2.62	0.00	1.00
## dem3_ethnicity*	48	608	2.89	0.85	3.00	2.93	0.00	1.00
## dem3_ethnicity_other*	49	608	14.83	1.28	15.00	15.00	0.00	1.00
## dem4_employment_status*	50	608	1.11	0.46	1.00	1.00	0.00	1.00
## dem5_occupation*	51	608	246.66	139.42	251.50	248.30	184.58	1.00

##	max	range	skew	kurtosis	se
## Response ID*	608.00	607.00	0.00	-1.21	7.12
## Group*	10.00	9.00	0.00	-1.21	0.12
## Duration (in seconds)	2065.00	2022.00	4.84	34.35	7.26
## Progress*	1.00	0.00	NaN	NaN	0.00
## User Language*	1.00	0.00	NaN	NaN	0.00
## Finished*	1.00	0.00	NaN	NaN	0.00
## Attention_check*	1.00	0.00	NaN	NaN	0.00
## timing_page_submit_LP	467.23	465.81	4.10	21.86	2.07
## click_count_LP	39.00	39.00	4.58	27.33	0.16
## item1_leader	7.00	6.00	-1.04	2.09	0.04
## timing_leader	208.16	206.52	7.36	75.18	0.60
## click_count_leader	13.00	13.00	2.25	6.01	0.08
## item2_independent	7.00	6.00	-1.07	1.47	0.05
## timing_independent	200.14	199.20	9.24	112.43	0.52
## click_count_independent	12.00	12.00	2.10	5.92	0.06
## item3_ambitious	7.00	6.00	-1.41	2.58	0.04
## timing_ambitious	257.49	256.50	8.72	87.63	0.77
## click_count_ambitious	8.00	8.00	1.75	3.36	0.05
## item4_loyal	7.00	6.00	-0.29	0.16	0.05
## timing_loyal	120.95	119.64	5.78	44.59	0.40
## click_count_loyal	12.00	12.00	2.36	7.89	0.06
## item5_sensitive	7.00	6.00	-0.19	-0.13	0.05
## timing_sensitive	1152.78	1151.55	19.37	421.91	2.06
## click_count_sensitive	10.00	10.00	1.92	4.42	0.06
## item6_warm	7.00	6.00	-0.36	-0.01	0.05
## timing_warm	169.46	168.28	7.85	74.59	0.51
## click_count_warm	12.00	12.00	2.28	7.63	0.06
## item7_compassionate	7.00	6.00	-0.22	0.03	0.05
## timing_compassionate	122.43	121.40	7.70	70.01	0.40
## click_count_compassionate	19.00	19.00	4.24	38.88	0.06
## item8_adaptable	7.00	6.00	-0.93	1.40	0.04
## timing_adaptable	106.76	105.73	6.48	49.22	0.39
## click_count_adaptable	14.00	14.00	2.82	16.29	0.05
## item9_sincere	7.00	6.00	-0.62	0.39	0.05
## timing_sincere	257.56	256.71	10.32	140.49	0.61
## click_count_sincere	8.00	8.00	1.80	3.45	0.05
## item10_reliable	7.00	5.00	-1.06	1.17	0.04
## timing_reliable	205.53	204.71	9.83	124.16	0.52
## click_count_reliable	10.00	10.00	2.00	5.74	0.05
## item11_truthful	7.00	6.00	-0.74	0.54	0.05
## timing_truthful	639.00	638.06	15.85	278.80	1.30
## click_count_truthful	8.00	8.00	1.60	2.86	0.05
## item12_race	-Inf	-Inf	NA	NA	NA
## timing_race	896.08	894.22	17.63	363.51	1.65
## click_count_race	25.00	25.00	4.20	28.80	0.10
## dem1_age	50.00	30.00	0.20	-1.01	0.33
## dem2_gender*	3.00	2.00	-0.63	-0.94	0.02
## dem3_ethnicity*	6.00	5.00	0.30	4.09	0.03
## dem3_ethnicity_other*	15.00	14.00	-8.21	70.33	0.05
## dem4_employment_status*	3.00	2.00	3.82	12.67	0.02
## dem5_occupation*	481.00	480.00	-0.05	-1.22	5.65

```
new_df <- new_df %>%
  mutate(masculinity = (item1_leader + item2_independent + item3_ambitious) / 3,
         femininity = (item4_loyal + item5_sensitive + item6_warm + item7_compassionate) / 4,
         neutral = (item8_adaptable + item9_sincere + item10_reliable + item11_truthful) / 4)

aggregated_scale <- new_df[, c("masculinity", "femininity", "neutral")]
alpha(aggregated_scale)
```

## Number of categories should be increased in order to count frequencies.

```
##
## Reliability analysis
## Call: alpha(x = aggregated_scale)
##
## raw_alpha std.alpha G6(smc) average_r S/N ase mean sd median_r
## 0.84 0.84 0.82 0.64 5.4 0.012 5.6 0.82 0.69
##
## 95% confidence boundaries
## lower alpha upper
## Feldt 0.82 0.84 0.86
## Duhachek 0.82 0.84 0.86
##
## Reliability if an item is dropped:
## raw_alpha std.alpha G6(smc) average_r S/N alpha se var.r med.r
## masculinity 0.86 0.86 0.76 0.76 6.3 0.011 NA 0.76
## femininity 0.82 0.82 0.69 0.69 4.5 0.015 NA 0.69
## neutral 0.65 0.65 0.48 0.48 1.9 0.028 NA 0.48
##
## Item statistics
## n raw.r std.r r.cor r.drop mean sd
## masculinity 608 0.82 0.83 0.70 0.62 5.9 0.90
## femininity 608 0.87 0.86 0.77 0.67 5.1 1.02
## neutral 608 0.93 0.94 0.91 0.85 5.7 0.89
```

*## all the aggregated scales highly correlate with the overall scale (highest corr neutral = 0.93)*  
*## so dropping neutral would have the most impact on alpha value, while dropping masculinity or femininity*

```
new_df <- new_df %>%
  mutate(total_response_time = timing_leader + timing_independent + timing_ambitious +
         timing_loyal + timing_sensitive + timing_warm + timing_compassionate +
         timing_adaptable + timing_sincere + timing_reliable + timing_truthful)
```

```
total_response_time_bygroup <- new_df %>%
  group_by(Group) %>%
  summarise(Mean = mean(total_response_time), Sd = sd(total_response_time))
total_response_time_bygroup
```

```
## # A tibble: 10 x 3
## Group Mean Sd
## <fct> <dbl> <dbl>
## 1 Asian_Man 82.5 73.8
## 2 Asian_Woman 77.8 58.0
```

```
## 3 Black_Man      82.7  86.1
## 4 Black_Woman  102.  114.
## 5 Blight_Man    76.2  66.0
## 6 Blight_Woman  73.2  39.7
## 7 Wasian_Man    87.5  67.3
## 8 Wasian_Woman  78.2  54.5
## 9 White_Man     95.1 156.
## 10 White_Woman  85.5  60.3
```

```
item_wise_response_time_bygroup <- new_df %>%
  group_by(Group) %>%
  summarise(Mean_item1_leader = mean(item1_leader), Sd_item1_leader = sd(item1_leader),
    Mean_item2_independent = mean(item2_independent), Sd_item2_independent = sd(item2_independent),
    Mean_item3_ambitious = mean(item3_ambitious), Sd_item3_ambitious = sd(item3_ambitious),
    Mean_item4_loyal = mean(item4_loyal), Sd_item4_loyal = sd(item4_loyal),
    Mean_item5_sensitive = mean(item5_sensitive), Sd_item5_sensitive = sd(item5_sensitive),
    Mean_item6_warm = mean(item6_warm), Sd_item6_warm = sd(item6_warm),
    Mean_item7_compassionate = mean(item7_compassionate), Sd_item7_compassionate = sd(item7_compassionate),
    Mean_item8_adaptable = mean(item8_adaptable), Sd_item8_adaptable = sd(item8_adaptable),
    Mean_item9_sincere = mean(item9_sincere), Sd_item9_sincere = sd(item9_sincere),
    Mean_item10_reliable = mean(item10_reliable), Sd_item10_reliable = sd(item10_reliable),
    Mean_item11_truthful = mean(item11_truthful), Sd_item11_truthful = sd(item11_truthful))
item_wise_response_time_bygroup
```

```
## # A tibble: 10 x 23
##   Group      Mean_item1_leader Sd_item1_leader Mean_item2_independent
##   <fct>          <dbl>          <dbl>          <dbl>
## 1 Asian_Man      5.58            1.24            5.81
## 2 Asian_Woman    5.6             1.06            5.58
## 3 Black_Man      5.57            1.35            5.7
## 4 Black_Woman    5.9             1.02            5.83
## 5 Blight_Man     6.02            0.852           5.90
## 6 Blight_Woman   5.82            0.840           5.85
## 7 Wasian_Man     5.94            0.965           5.90
## 8 Wasian_Woman   5.63            0.938           5.83
## 9 White_Man      5.59            1.04            5.76
## 10 White_Woman   5.72            1.03            5.92
## # i 19 more variables: Sd_item2_independent <dbl>, Mean_item3_ambitious <dbl>,
## #   Sd_item3_ambitious <dbl>, Mean_item4_loyal <dbl>, Sd_item4_loyal <dbl>,
## #   Mean_item5_sensitive <dbl>, Sd_item5_sensitive <dbl>,
## #   Mean_item6_warm <dbl>, Sd_item6_warm <dbl>, Mean_item7_compassionate <dbl>,
## #   Sd_item7_compassionate <dbl>, Mean_item8_adaptable <dbl>,
## #   Sd_item8_adaptable <dbl>, Mean_item9_sincere <dbl>, Sd_item9_sincere <dbl>,
## #   Mean_item10_reliable <dbl>, Sd_item10_reliable <dbl>, ...
```